The Expected Peak-to-Average Power Ratio of White Gaussian Noise in Sampled I/Q Data

Adam Wunderlich and Aric Sanders

Abstract—One of the fundamental endeavors in radio frequency (RF) metrology is to measure the power of signals, where a common aim is to estimate the peak-to-average power ratio (PAPR), which quantifies the ratio of the maximum (peak) to the mean value. For a finite number of discrete-time samples of baseband in-phase and quadrature (I/Q) white Gaussian noise (WGN) that are independent and identically distributed with zero mean, we derive a closed-form, exact formula for mean PAPR that is well-approximated by the natural logarithm of the number of samples plus Euler's constant. Additionally, we give related theoretical results for the mean crest factor (CF). After comparing our main result to previously published approximate formulas, we examine how violations of the WGN assumptions in sampled I/Q data result in deviations from the expected value of PAPR. Finally, utilizing a measured RF I/Q acquisition, we illustrate how our formula for mean PAPR can be applied to spectral analysis with spectrograms to verify when measured RF emissions are WGN in a given frequency band.

Index Terms—Crest factor (CF), max hold, peak detection, peak-to-average power ratio (PAPR), Rayleigh distribution.

I. INTRODUCTION

IN RADIO frequency (RF) electronic measurements, it is common to perform peak detection, where the maximum amplitude or power of data collected over a given time interval is recorded. In particular, the peak-to-average power ratio (PAPR) and crest factor (CF), equal to the square root of PAPR, arise in the design of communication signals and power amplifiers for RF transmitters [1], [2], [3]. While it is important to understand the PAPR and CF for transmitted continuous-time communication signals and their impact on power amplifiers, it is equally important to characterize PAPR and CF for sampled RF measurements of received signals, e.g., with a signal analyzer. Notable applications of received signals include spectrum monitoring [4], [5] and the detection of transient electromagnetic interference (EMI) [6]. The appropriate selection of settings for measurement instrumentation in these domains requires foreknowledge of the peak level that can be expected over a given time period.

Modulated RF signals are typically measured, processed, and stored in an equivalent baseband representation, called the in-phase and quadrature (I/Q) signal representation [7]. Furthermore, the noise level in the absence of a signal is typically dominated by thermal noise in the in-phase (I) and quadrature (Q) components that can be modeled as a stationary, zero-mean, white Gaussian process [8], [9]. Therefore, the expected PAPR of stationary white Gaussian noise (WGN) in I/Q data is of high interest to differentiate thermal noise from other noise and signal types.

The difference between PAPR assessment for continuous-time and discrete-time measurements is often ignored [10], because the Nyquist-Shannon sampling theorem implies that a band-limited signal may be perfectly reconstructed from samples if the sampling rate is greater than or equal to the Nyquist rate. However, real signals are not perfectly band-limited, and the analog-to-digital (A/D) and digital-to-analog (D/A) conversion process is non-ideal [11]. Consequently, the PAPR of a discrete-time sample is not identical to the PAPR for the corresponding continuous-time signal. Bounds on the difference between the peak of a continuous-time, band-limited signal and the peak value of a discrete-time, over-sampled measurement were studied in [12] and [13].

Extrema theory for continuous-time, stationary random processes is reviewed in [14] and [15]. In particular, notable results concerning the expected number of local peaks above a specified level per unit time were developed for band-limited, continuous-time signals in [16], [17], [18]. This theory was later applied by Ochai and Imai [19] to derive accurate expressions for the PAPR distribution of band-limited, continuous-time orthogonal frequency-division multiplexing (OFDM) communication signals.

The present work concerns the expected PAPR for discrete-time, sampled I/Q measurements of WGN, which is relevant to digital signal analyzer measurements. Namely, using extreme value theory [20] and the theory of order statistics [21], we develop highly accurate, closed-form formulas for the mean PAPR and mean CF for a finite collection of n independent, identically distributed (i.i.d.) I/Q samples of stationary, zero-mean WGN. To our knowledge, the theoretical results given here and their connection to measurement have not been presented in the literature. For this problem, approximate formulas for the mean PAPR were presented in [3] and [22]. Also, exact cumulative distribution functions (CDFs) of PAPR and CF for WGN I/Q samples were given in [19], but not analyzed further. Compared to these prior works, we present additional theoretical insights, context, and applications. Namely, we examine how non-ideal I/Q measurements of thermal noise can

violate the WGN assumptions and impact the expected PAPR. Further, we show how the expected mean PAPR formula can be applied to spectrograms, which are widely used to visualize the frequency content of a nonstationary signal as it varies in time, to verify when measured RF emissions in a given frequency band consist of thermal noise.

II. THEORETICAL RESULTS

Consider a baseband signal, X, with I/Q signal components, X_I and X_Q. Denote the amplitude (or envelope) of X as $A = \sqrt{|X_I|^2 + |X_Q|^2}$ and the instantaneous power (or squared amplitude)[1] as $P = |X_I|^2 + |X_Q|^2$. Unless otherwise stated, we assume that the I/Q components can be modeled as WGN. Specifically, X_I and X_Q, are taken to be i.i.d. normal random variables with zero mean and variance σ^2, denoted $X_I \sim \mathcal{N}(0, \sigma^2)$ and $X_Q \sim \mathcal{N}(0, \sigma^2)$.

We denote sample numbers with subscripts, and order statistics with subscripts in parentheses. For example, given a sample of n power measurements $P_1, P_2, \ldots, P_n$, the order statistics are $P_{(1)}, P_{(2)}, \ldots, P_{(n)}$, where $P_{(r)}$ is the rth largest value. We write $P_{\max} = P_{(n)}$ to denote the maximum value. Additionally, for a random variable, X, we denote its CDF as $F_X(x)$, its probability density function (PDF) as $f_X(x)$, and the expected value as $\mathrm{E}[X]$.

For a given set of power measurements, the PAPR is defined as the peak power value divided by the mean power, i.e., $\mathrm{PAPR} = P_{\max}/\mathrm{E}[P]$. Similarly, the CF is defined as $\mathrm{CF} = A_{\max}/\sqrt{\mathrm{E}[P]}$, where $A_{\max} = \sqrt{P_{\max}}$, i.e., $\mathrm{PAPR} = \mathrm{CF}^2$. Below, it will be convenient to utilize scaled versions of P and A defined as $\widetilde{P} = P/\mathrm{E}[P]$ and $\widetilde{A} = A/\sqrt{\mathrm{E}[P]}$ so that $\mathrm{PAPR} = \widetilde{P}_{\max}$ and $\mathrm{CF} = \widetilde{A}_{\max}$.

We begin with a review of the established distribution theory for PAPR and CF. To the best of our knowledge, the subsequent theoretical results are new unless stated otherwise.

A. Distribution Theory

Since $X_I \sim \mathcal{N}(0, \sigma^2)$ and $X_Q \sim \mathcal{N}(0, \sigma^2)$, it follows that the amplitude, A, has a Rayleigh distribution with scale parameter, σ, and the instaneous power, P, has an exponential distribution with rate parameter, $\lambda = 1/(2\sigma^2)$ [23], [24]. Here, we indicate this with the notation $A \sim \mathrm{Rayleigh}(\sigma)$ and $P \sim \mathrm{Exp}(1/(2\sigma^2))$, where the quantity in parentheses denotes the parameter value.

For $x \geq 0$, the CDFs for A and P are $F_A(x) = 1 - \exp[-x^2/(2\sigma^2)]$ and $F_P(x) = 1 - \exp[-\lambda x] = 1 - \exp[-x/(2\sigma^2)]$, and the mean values are $\mathrm{E}[A] = \sigma\sqrt{\pi/2}$ and $\mathrm{E}[P] = 1/\lambda = 2\sigma^2$ [24]. Applying the CDF transformation theorem for an increasing function of a random variable [25, Theorem 2.1.3], it follows that for $x \geq 0$, the CDFs of $\widetilde{A} = A/(\sigma\sqrt{2})$ and $\widetilde{P} = P/(2\sigma^2)$ are $F_{\widetilde{A}}(x) = 1 - \exp[-x^2]$ and $F_{\widetilde{P}}(x) = 1 - \exp[-x]$. In terms of the previously introduced distribution notation, $\widetilde{A} \sim \mathrm{Rayleigh}(1)$ and $\widetilde{P} \sim \mathrm{Exp}(1)$, i.e., $\sigma = 1$ and $\lambda = 1$ for the Rayleigh and exponential distributions, respectively.

[1]For generality, we avoid assuming physical units and the requisite scale factors necessary to convert between squared amplitude and power units.

Suppose we have n i.i.d. measurements of $X_I \sim \mathcal{N}(0, \sigma^2)$ and $X_Q \sim \mathcal{N}(0, \sigma^2)$, yielding n i.i.d. measurements, $\widetilde{P}_1, \widetilde{P}_2, \ldots, \widetilde{P}_n$, of $\widetilde{P} \sim \mathrm{Exp}(1)$ and n i.i.d. measurements, $\widetilde{A}_1, \widetilde{A}_2, \ldots, \widetilde{A}_n$, of $\widetilde{A} \sim \mathrm{Rayleigh}(1)$. By a well-known formula for the CDF of the maximum of n i.i.d. samples from a known CDF [21, Eq. 2.1.1], the CDFs of PAPR and CF are

$$F_{\mathrm{PAPR}}(x) = (1 - \exp[-x])^n \tag{1}$$

and

$$F_{\mathrm{CF}}(x) = (1 - \exp[-x^2])^n \tag{2}$$

for $x \geq 0$. Note that the distributions of PAPR and CF do not depend on σ, i.e., they are independent of the mean power. The CDFs (1) and (2) were previously given in [19].

From the CDFS it follows that the quantile functions (inverse CDFs) of PAPR and CF are

$$Q_{\mathrm{PAPR}}(p) = -\ln(1 - p^{1/n}) \tag{3}$$

and

$$Q_{\mathrm{CF}}(p) = \sqrt{-\ln(1 - p^{1/n})} \tag{4}$$

for $0 \leq p < 1$. The above expressions are useful for engineering design when it is desired to know specific percentiles of the distributions.

Also, differentiating the CDFs yields corresponding expressions for the respective PDFs. Namely, the PDF of PAPR is

$$f_{\mathrm{PAPR}}(x) = n \exp[-x](1 - \exp[-x])^{n-1} \tag{5}$$

To express the PAPR PDF on a decibel (dB) scale, we apply the PDF transformation theorem [25, Theorem 2.1.5] with $\mathrm{PAPR}_{\mathrm{dB}} = 10 \log_{10}(\mathrm{PAPR})$ to obtain

$$f_{\mathrm{PAPR}_{\mathrm{dB}}}(y) = \frac{n}{10} \ln(10) 10^{\frac{y}{10}} e^{-10^{y/10}} \left(1 - e^{-10^{y/10}}\right)^{n-1}. \tag{6}$$

B. Asymptotic Distributions

From extreme value theory [20], it is known that the distribution for the maximum of an i.i.d. sample from an exponential or Rayleigh distribution asymptotically approaches a Gumbel distribution [20, Sec. 5.2] as the sample size increases. Specifically, letting $G(z) = \exp[-\exp(-z)]$ denote the standard Gumbel CDF, $\lim_{n \to \infty} F_{\mathrm{PAPR}}(\ln(n) + z) = G(z)$ and $\lim_{n \to \infty} F_{\mathrm{CF}}(\sqrt{\ln n} + z/(2\sqrt{\ln n})) = G(z)$ [20, p. 105-108]. Hence, for large n

$$F_{\mathrm{PAPR}}(x) \approx \exp[-\exp(-(x - \ln n))] \tag{7}$$

and

$$F_{\mathrm{CF}}(x) \approx \exp[-\exp(-(x - \sqrt{\ln n})2\sqrt{\ln n})]. \tag{8}$$

The above CDFs take the form of a general Gumbel distribution, which has CDF $F(x; a, b) = \exp[-\exp(-(x - a)/b)]$ and mean $a + \gamma b$, where $\gamma = 0.5772156649\ldots$ is Euler's constant [20, p. 185].

C. Mean Peak-to-Average Power Ratio

First, we obtain an approximate asymptotic expression for the mean PAPR from (7) and the formula for the mean of the Gumbel distribution. Namely, for large n

$$\mathrm{E[PAPR]} \approx \ln n + \gamma. \tag{9}$$

An exact closed-form formula for the mean PAPR is derived from established results on order statistics for the standard exponential distribution, which imply that the mean of the rth order statistic, $\widetilde{P}_{(r)}$, is [24, Eq. 19.12]

$$\mathrm{E}[\widetilde{P}_{(r)}] = \sum_{j=1}^{r} \frac{1}{n-j+1} = \sum_{k=n-r+1}^{n} \frac{1}{k}. \tag{10}$$

Setting $r = n$ for the maximum value, $\widetilde{P}_{\max} = \mathrm{PAPR}$, gives

$$\mathrm{E[PAPR]} = H_n \tag{11}$$

where $H_n = \sum_{k=1}^{n} 1/k$ denotes the nth harmonic number. The above formula is our main result and will be applied later in Sections IV and V.

Efficient computation of H_n for large n can be carried out with the asymptotic approximation [26, p. 480]

$$H_n \approx \ln n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4}. \tag{12}$$

The first two terms, which correspond to (9), are sufficient to ensure that the relative approximation error is less than 0.1% when $n \geq 100$.

D. Mean Crest Factor

First, from (8) and the formula for the mean of a Gumbel distribution, it follows that for large n, the mean CF is roughly

$$\mathrm{E[CF]} \approx \sqrt{\ln n} + \gamma/(2\sqrt{\ln n}). \tag{13}$$

The above approximation was found in [27, Eq. (59)] by other means in the context of narrow-band sea wave maxima.

We can also use (11) to derive an upper bound for the mean CF. Since $\mathrm{CF} = \sqrt{\mathrm{PAPR}}$, and because square root is a concave function, Jensen's inequality [25, Thm. 4.7.7] implies that $\mathrm{E[CF]} \leq \sqrt{\mathrm{E[PAPR]}}$. Thus, (11) yields

$$\mathrm{E[CF]} \leq \sqrt{H_n}. \tag{14}$$

An exact closed-form expression for the mean CF can be obtained from known results for the mean order statistics of a Rayleigh distributed sample. Namely, starting with [24, Eq. 18.82] and simplifying yields

$$\mathrm{E[CF]} = \frac{\sqrt{\pi}}{2} \sum_{k=1}^{n} \binom{n}{k} (-1)^{k-1} k^{-1/2} \tag{15}$$

where $\binom{n}{k}$ denotes the binomial coefficient. Unfortunately, the above formula is not useful for computation when $n > 50$ (approximately), due to numerical error.

Instead, we express the mean CF as an integral that can be accurately evaluated for large n with standard numerical quadrature routines. The probability integral transformation [25, Th. 2.1.10] implies that the random variable obtained by transforming CF with its CDF, $U = F_{\mathrm{CF}}(\mathrm{CF})$, follows a standard uniform distribution on the interval $[0, 1]$. Thus, transforming a standard uniform random variable, U, with the CF quantile function (inverse CDF) yields $Q_{\mathrm{CF}}(U) = \mathrm{CF}$, and hence, $\mathrm{E[CF]} = \mathrm{E}[Q_{\mathrm{CF}}(U)]$. Writing the last expectation in integral form with (4) gives

$$\mathrm{E[CF]} = \int_0^1 \sqrt{-\ln(1 - u^{1/n})}\, du. \tag{16}$$

Fig. 1a shows the relative error of the approximation (13) and the bound (14) (the lower plot, Fig. 1b, is discussed in the next section). The relative error, in percent, is calculated as Rel Err $= (\mathrm{approx} - \mathrm{exact})/\mathrm{exact} \times 100$, where approx denotes the values from (13) and (14) and exact denotes the value from (16), all expressed on a linear scale. The definite integral in (16) was evaluated using the 'integrate.quad' method in the SciPy python library [28][2] with the relative error tolerance parameter set to 10^{-10}. From this plot, we see that (13) and (14) are excellent approximations for the mean CF, with (14) being slightly better. Thus, (13) and (14) are a simple, accurate alternatives to numerically evaluating the integral in (16).

III. COMPARISON WITH PREVIOUSLY PUBLISHED APPROXIMATE FORMULAS FOR MEAN PAPR

We compare the exact result (11) presented in Section II-C to two approximate formulas from the engineering literature for the mean PAPR of WGN in I/Q data. The first formula, given in a book on microwave measurements [3, p. 588], states that for noise-like signals, the expected PAPR, expressed in linear units, is roughly

$$\mathrm{E[PAPR]} \approx \ln(n) \tag{17}$$

where n is the number of samples. This formula is equivalent to the first term of the asymptotic expansion (12). Thus, it underestimates the expected PAPR, with the difference approaching $\gamma \approx 0.5772$ as n approaches infinity.

An equipment vendor application note gives the following formula, expressed in linear units, for the average PAPR found from "a combination of analysis, approximation and experimentation" [22, p. 17-18]

$$\mathrm{E[PAPR]} \approx \ln(2\pi\tau\mathrm{BW}_i + e) \tag{18}$$

where $e = 2.71828\ldots$ is Euler's number, τ is the observation period and BW_i is the one-sided impulse bandwidth for the resolution bandwidth (RBW) filter. Ref. [4, p. 7] applies the above formula in the context of spectrum monitoring.

To express Eq. (18) in terms of the number of samples, we assume that the one-sided impulse bandwidth is well approximated by one half the sampling rate, f_s, i.e., the signal is critically sampled.[3] Multiplying by the observation period, then yields $\tau\mathrm{BW}_i \approx \tau f_s/2 = n/2$, where n is the number of samples. Hence, we obtain

$$\mathrm{E[PAPR]} \approx \ln(\pi n + e). \tag{19}$$

---

[2]Certain equipment, instruments, software, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement of any product or service by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

[3]In practice, an anti-aliasing filter may reduce the bandwidth further.
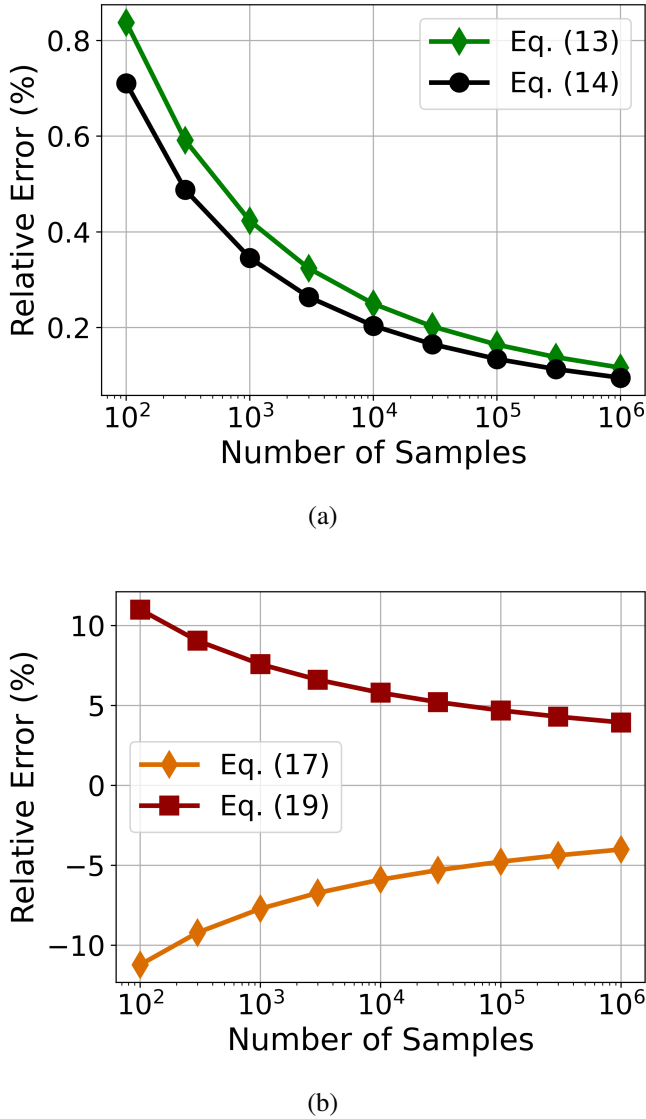
(a)

(b)

Fig. 1. Relative error in approximations for mean CF and mean PAPR with WGN I/Q data. (a) Mean CF approximations (see Section II-D). (b) Mean PAPR approximations (see Section III).

In order to compare the above formula to (11), observe that since $\pi n + e = \pi n(1 + e/(\pi n))$,

$$\ln(\pi n + e) = \ln(\pi n) + \ln\left(1 + \frac{e}{\pi n}\right) \qquad (20)$$

$$\approx \ln n + \ln \pi + \frac{e}{\pi n} \qquad (21)$$

where the last approximation follows from the first-order Maclaurin expansion for $\ln(1+x)$. Finally, since the last term approaches zero as n goes to infinity, we find that (19) is roughly $\mathrm{E}[\mathrm{PAPR}] \approx \ln n + \ln \pi$ for large n. Thus, compared with the asymptotic expansion (12), (19) overestimates the mean PAPR, with the error approaching $\ln \pi - \gamma \approx 0.5675$ as n approaches infinity.

Fig. 1b shows the relative difference, in percent, between the approximate expressions (17) and (19) and the exact result (11). These plots indicate that the approximate formulas result in relative errors ranging from greater than 10% for $n = 100$ to less than 5% for $n > 10^5$. As expected,

(17) and (19) underestimate and overestimate the true value, respectively. Even on a decibel scale, these errors can be significant for some applications. For example, for $n = 1000$, the approximate formulas yield errors of roughly $\pm 0.3\,\mathrm{dB}$.

IV. EXAMPLES OF WGN ASSUMPTION VIOLATIONS

Our theoretical findings are built upon assumptions about the nature of the noise and the operational space of the instruments recording it. Namely, our results require that the recorded samples of I and Q components are zero-mean WGN, implying that the I and Q components satisfy the following conditions:

(1) not dominated by nonlinear distortion effects, such as quantization noise

(2 balanced, i.e., the gains of each RF leg are equal, and the phase difference is $90°$

(3) statistically independent, white Gaussian processes.

Below, we provide examples illustrating how breakdowns in the above assumptions can occur in practical measurements of thermal noise and impact the expected PAPR. Conversely, when measuring thermal noise, significant deviations from the expected mean PAPR for WGN provide evidence that I/Q measurements are non-ideal and violate one or more of the above assumptions.

A. Nonlinear Distortion

To illustrate the impact of nonlinear distortions arising from improper selection of reference level (the maximum expected input power), we collected repeated RF measurements with a National Instruments PXIe-5646 vector signal transceiver (VST), which has a 14-bit analog-to-digital converter (ADC) resolution [29]. The RF input to the VST was terminated with a matched, $50\,\Omega$ load so that the input signal consisted of room-temperature thermal noise. Repeated captures of I/Q samples were collected with a center frequency of $5\,\mathrm{GHz}$ at a sampling rate of 30 Megasamples per second (MS/s) with four different VST reference level settings: $-50\,\mathrm{dBm}$, $10\,\mathrm{dBm}$, $20\,\mathrm{dBm}$, and $30\,\mathrm{dBm}$. The $-50\,\mathrm{dBm}$ reference level was sufficiently low so as to avoid quantization effects, whereas the other reference levels were sufficiently large to ensure impacts from nonlinear quantization.

Fig. 2 (top) plots the mean PAPR estimated from 120 repeated measurements at each sample size for each reference level setting, as well as the theoretical value for I/Q WGN predicted by (11). Error bars are suppressed, since in all cases 95% confidence intervals were smaller than $\pm 0.19\,\mathrm{dB}$. From this plot, we see that when the VST reference level was set to -50 dBm, the empirical mean PAPR was very close to the theoretical mean value for WGN. By contrast, when the VST reference level was set to the higher levels, the recorded noise was influenced by nonlinear distortion stemming from quantization effects and the resulting mean PAPR was lower than for WGN. Fig. 2 (bottom) plots empirical PDFs of PAPR for the $n = 10^4$ case, estimated with the nonparametric kernel density estimation (KDE) method and compares them to the theoretical PDF given by (6). KDE was implemented using 'stats.gaussian_kde' in the SciPy Python library [28]
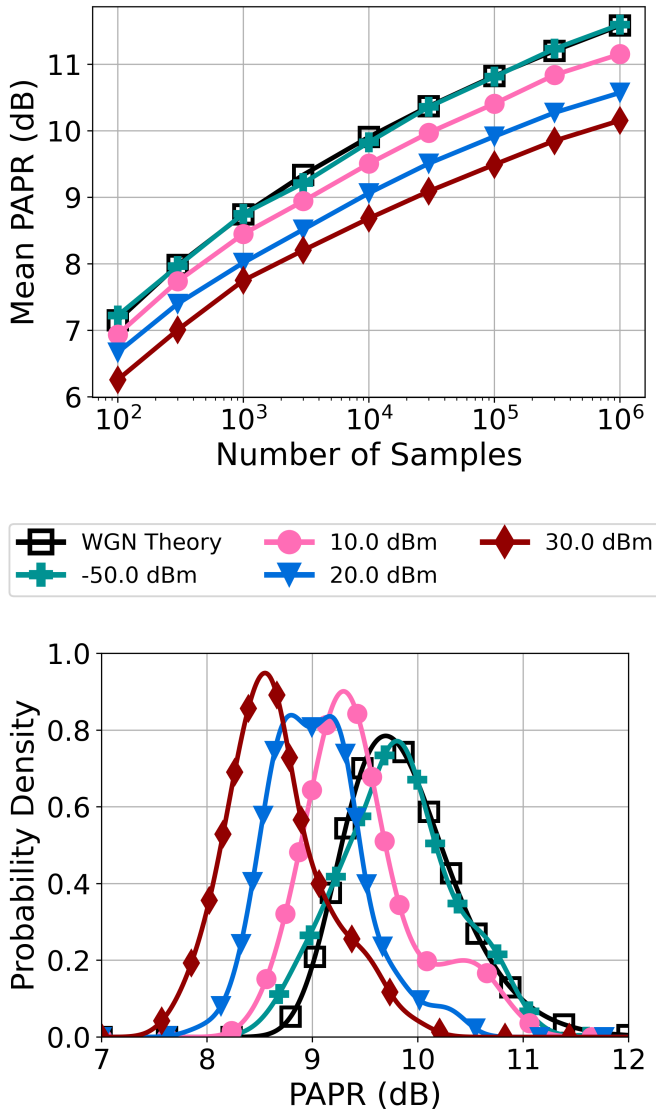
Fig. 2. Vector signal transceiver (VST) measurements comparing the PAPR of I/Q thermal noise measurements with various reference levels. Top: Mean PAPR vs sample size. Bottom: Estimated PDFs for $n = 10^4$ samples.

with default settings, which apply Gaussian kernels with the bandwidth parameter set with Scott's rule [30]. As expected, the empirical PDF for the -50 dBm reference level had excellent agreement with the theoretical PDF (6).

B. I/Q Imbalance

In RF transmitters and receivers, I/Q imbalance (or mismatch) results from gain and phase mismatches on the I and Q branches [31, Sec. 2.7]. Let $\Delta\phi$ be the phase mismatch between the I and Q components, and let $\Delta_G = |G_I - G_Q|/G_Q$ be the relative gain mismatch, where G_I and G_Q are the gains for the I and Q branches. Also, let $A = 1 - \Delta G/2$ and $B = 1 + \Delta G/2$. Using matrix multiplication, I/Q imbalance in an RF receiver can be modeled as [31, Eq. (2.131)]

$$\begin{pmatrix} Y_I \\ Y_Q \end{pmatrix} = \begin{pmatrix} A\cos(\Delta\phi/2) & A\sin(\Delta\phi/2) \\ B\sin(\Delta\phi/2) & B\cos(\Delta\phi/2) \end{pmatrix} \begin{pmatrix} X_I \\ X_Q \end{pmatrix} \quad (22)$$

where X_I and X_Q are the I/Q components of the ideal (balanced) baseband signal, and Y_I and Y_Q are the I/Q components of the baseband signal with I/Q imbalance.

Monte Carlo simulations were performed to demonstrate the impacts of I/Q imbalance on the expected PAPR. For each sample size, I/Q samples were independently drawn from standard normal $\mathcal{N}(0, 1)$ distributions and transformed with (22) for $(\Delta G, \Delta\phi) = (0.05, 10°), (0.1, 15°)$, and $(0.2, 20°)$. In addition, the case of no I/Q imbalance, $(\Delta G, \Delta\phi) = (0, 0)$, was simulated for baseline comparison. Mean PAPR estimated from 10^4 independent Monte Carlo trials for each sample size is plotted in Fig. 3 (top) together with the theoretical value (11) for ideal I/Q WGN. In all cases, the 95% confidence intervals were smaller than ± 0.022 dB. From this plot, we observe that the expected PAPR increased as the I/Q imbalance became more severe. For example, in the $(\Delta G, \Delta\phi) = (0.2, 20°)$ case with $n = 10^4$ samples, the mean PAPR increased by roughly 0.8 dB over the baseline case with no I/Q imbalance. Fig. 3 (bottom) plots PDFs for the $n = 10^4$ case estimated with the KDE method described in Section IV-A and compares them with the theoretical PDF for ideal I/Q WGN given by (6). For both plots in Fig. 3 in the ideal case with no I/Q imbalance, there was excellent agreement between the Monte Carlo results and theory.

C. Correlated Samples

Finally, we examined the impacts of I/Q measurement correlations on the expected PAPR. Specifically, Monte Carlo simulations were performed to assess PAPR of correlated samples generated by low-pass filtering I/Q WGN. This example gives insight into the case of band-limited Gaussian noise when the (one-sided) bandwidth is smaller than the Nyquist frequency, equal to one half the sampling rate.

For each sample size, I/Q samples were independently drawn from standard normal $\mathcal{N}(0, 1)$ distributions and then filtered with a 40th-order (41 tap) low-pass digital finite impulse response (FIR) filter. The digital FIR filter was designed using the window method with a Hamming window [11], as implemented by the 'signal.firwin' method in the SciPy python library [28] with default parameters. To avoid transients at the beginning or end of the length n filtered noise sequence, an additional number of WGN I/Q samples equal to the filter order were added to both ends of the sequence and then discarded after filtering. The filter cutoff frequency, f_c, specified in normalized digital frequency[4] units of cycles per sample, took the values $\{0.025, 0.05, 0.1, 0.25\}$. Frequency responses for the lowpass FIR filters are shown in Fig. 4.

Mean PAPR estimated from 10^4 independent Monte Carlo trials for each sample size is plotted in Fig. 5 (top) together with the theoretical mean PAPR (11) for ideal I/Q WGN. In all cases, 95% confidence intervals were smaller than ± 0.025 dB. We see that the mean PAPR generally decreased as the filter cutoff, f_c, decreased. Moreover, the mean PAPR was very close to the theoretical WGN value for $f_c \geq 0.25$, indicating that (11) is an accurate approximation when $0.25 \leq f_c \leq 0.5$.

---

[4]In terms of analog frequency, f, and sampling rate, f_s, the normalized digital frequency is f/f_s and lies in the interval $[0, 0.5]$.
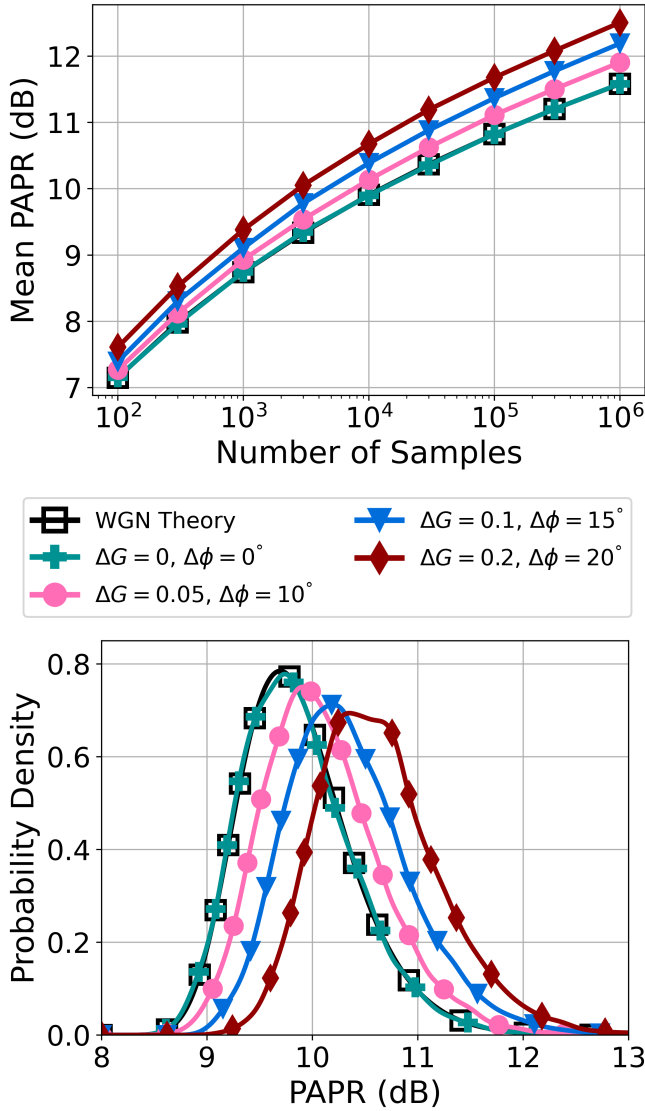
Fig. 3. Simulated I/Q Imbalance Example. Top: Mean PAPR vs. sample size. Bottom: Estimated PDFs for $n = 10^4$ samples.

Fig. 5 (bottom) compares PDFs estimated with the KDE method described in Section IV-A for the $n = 10^4$ sample size to the I/Q WGN theoretical PDF (6).

V. EXAMPLE APPLICATION TO SPECTRAL ANALYSIS

Finally, we demonstrate how our main result, (11), can be applied to spectral analysis of sampled I/Q signals to determine if a given frequency band contains only WGN. Specifically, we illustrate the application to spectrograms, which are widely used to assess the frequency content of a nonstationary signal as it varies in time. A spectrogram is the magnitude squared of a sampled short-time Fourier transform (STFT), computed by applying the discrete Fourier transform (DFT) successively to shifted, windowed signal segments [11, Sec. 15.3].

The application relies on the fact that the real and imaginary components of the DFT of a windowed realization of a stationary, circularly symmetric complex WGN process are
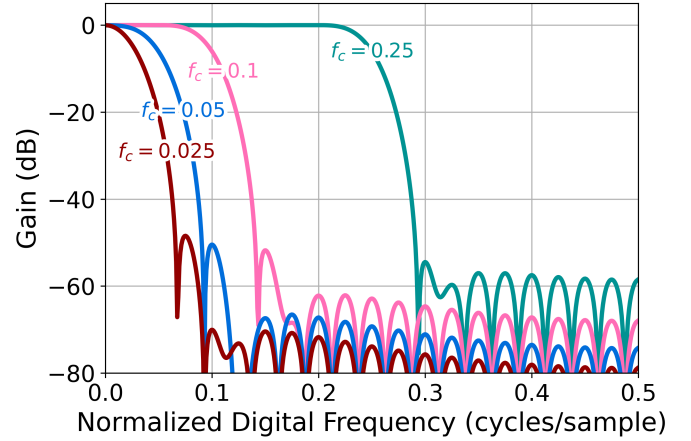

Fig. 4. Frequency responses of low-pass digital filters used for correlated noise simulations. The cutoff frequency, f_c, is specified in cycles per sample.

i.i.d. zero-mean WGN [32]. In particular, suppose that an N-point discrete-time random process, $y[t]$, can be decomposed as the sum of (a) a random process only supported on a proper subset, B, of the discrete-frequency DFT domain and (b) a stationary, circularly symmetric complex WGN process. Using the linearity of the DFT, it then follows that the real and imaginary components of the DFT of $y[t]$ are i.i.d. zero-mean WGN on the complement of B. Hence, the mean PAPR of the spectrogram of $y[t]$ for each frequency bin not in B is given by (11) with n equal to the number of spectrogram time bins.

To illustrate how the above theoretical observation can be applied, we used a Deepwave Digital AIR-T 7201B [33] software defined radio (SDR) to record I/Q data at a sample rate of 30 Ms/s, for 30 ms at a center frequency of 3700 MHz. A general-purpose antenna, band-pass filter, and low-noise amplifier were used as an RF front-end prior to the SDR to amplify signals in the 3 GHz to 4 GHz frequency range. The amplitude measured by the SDR was not calibrated. The acquisition band, which spanned 3685 MHz to 3715 MHz, covers the transition between the 3550 MHz to 3700 MHz Citizens Broadband Radio Service (CBRS) band [34] and the 3700 MHz to 3980 MHz band [35], which are both used for mobile wireless communications in the United States.

A spectrogram computed from the I/Q capture is shown in Fig. 6 (top), with the frequency range cropped to 3690 MHz to 3710 MHz for this example. The spectrogram was calculated with the 'signal.ShortTimeFFT.spectrogram' method in the Python SciPy library [28], using a Hann window of length 512 and 50% window overlap, yielding a total of 512 frequency bins and 3517 time bins. To suppress the SDR local oscillator (LO) signal at 3700 MHz, the 'constant' detrending option was applied, which subtracted the mean from each window segment. The spectrogram, which visualizes the (uncalibrated) signal power across time and frequency shows visible mobile wireless emissions in the 3700 MHz to 3710 MHz band, but no obvious emissions in the 3690 MHz to 3700 band.

PAPR was calculated in each spectrogram frequency bin by finding the ratio of the maximum to the mean across all 3517 time bins and is plotted in Fig. 6 (bottom) on a
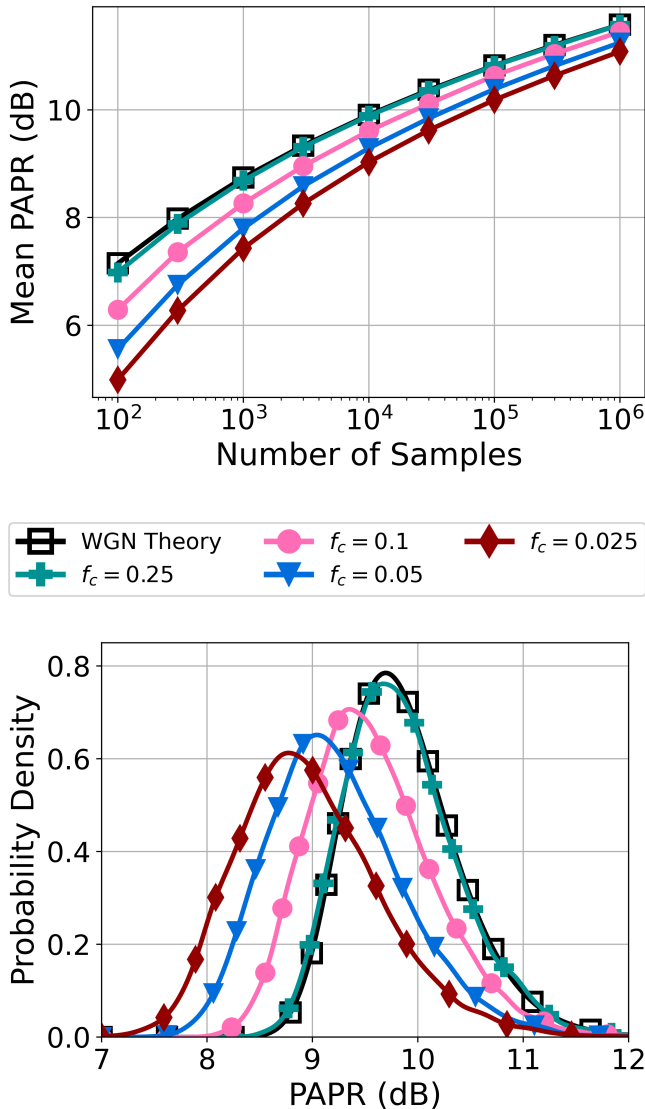
Fig. 5. Correlated noise simulations with low-pass filter cutoff frequency, f_c, specified in cycles per sample. (Top) Mean PAPR versus sample size. (Bottom) Estimated PDFs for $n = 10^4$ samples.

Fig. 6. Spectrogram example for an I/Q capture centered at 3.7 GHz. Top: Spectrogram with grayscale spanning -110 dB to -60 dB (uncalibrated power units). Bottom: PAPR vs. frequency.

decibel scale. The mean PAPR in the 3690 MHz to 3700 MHz band was estimated to be 9.48 dB (95% confidence interval $[9.38, 9.58]$ dB) by averaging PAPR across the 170 frequency bins covering that band. The mean PAPR given by (11) with $n = 3517$, was found to be 9.42 dB, as indicated by the thick dashed line in Fig. 6 (bottom). There was strong agreement between the observed mean PAPR in the 3690 MHz to 3700 MHz band and the theoretical mean value for WGN, supporting the hypothesis that the band contained only thermal noise during the observation interval. By contrast, the observed PAPR in the 3700 MHz to 3710 MHz band was significantly higher, consistent with mobile wireless emissions.

This example illustrates how the theoretical formula (11) can be used as a quick, simple method to determine if a given frequency band contains only thermal noise, which is useful for signal detection and channel occupancy evaluations. We make three observations. First, the approach is robust to the
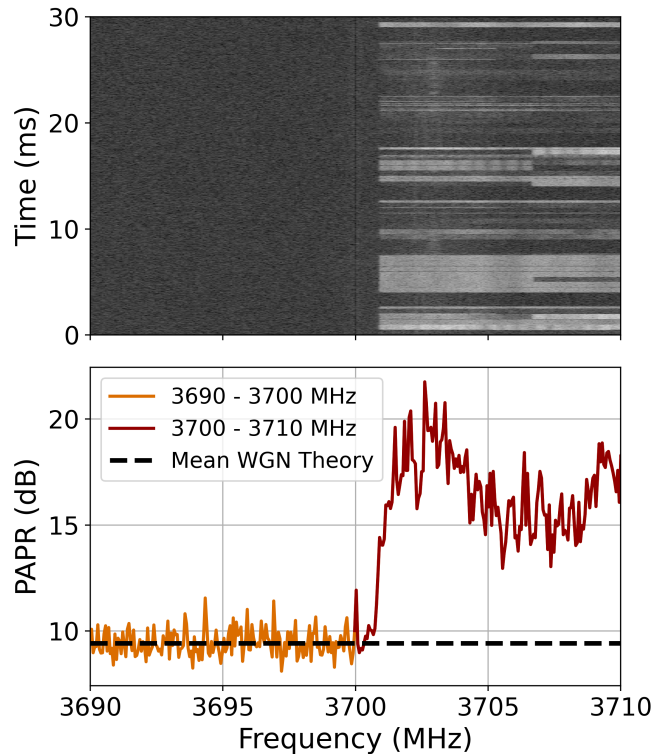
choice of spectrogram window function. Namely, the effect of windowing cancels out in the PAPR since in the case of WGN, the value of each spectrogram frequency bin follows an exponential distribution with the distribution parameter scaled by the energy in the window sequence [32, Eq. 44]. Second, we have found this method to be fairly robust to the degree of spectrogram window overlap, e.g., 50%, 75%, although extreme degrees of windows overlap ($> 90\%$) induce too much statistical dependence between window segments, rendering (11) inaccurate. Last, since PAPR is generally independent of power calibration factors, this method does not require calibrated I/Q values.

VI. DISCUSSION & CONCLUSIONS

After reviewing established distribution theory for PAPR and CF of zero-mean, i.i.d. WGN in sampled I/Q data, we presented approximate and exact theoretical expressions for the mean PAPR and mean CF. In particular, we derived an exact, closed-form formula for the mean PAPR, (11), and compared it to approximate formulas given in the instrumentation literature, demonstrating that the previously published formulas are asymptotically biased. Additionally, we used (11) to derive a closed-form upper bound to the mean CF and showed that it is a very accurate approximation to the mean CF. Next, we examined circumstances where violations of the WGN assumptions may arise in practical I/Q measurements of thermal noise and lead to deviations from the expected PAPR. Consequently, when measuring thermal noise, signifi-

cant deviations from the expected PAPR for I/Q WGN indicate potential non-ideal conditions. Finally, we presented a real-world example with a spectrogram showing how the mean PAPR formula provides a simple, quick method to check if observations of a given frequency band consist of only thermal noise.

Our theoretical development was based on the assumption of white, uncorrelated I/Q samples. However, our evaluations of low-pass filtered noise (see Section IV-C) and a spectrogram with overlapping windows (see Section V) indicated that the expected mean PAPR is somewhat robust to data correlations. Specifically, for low-pass filtered noise with digital filter cutoff frequencies greater than 0.25 cycles per sample, the mean PAPR was very close to the theoretical mean PAPR for uncorrelated white noise. Additionally, for a measured spectrogram, we found close agreement between the mean PAPR for thermal noise and the theoretical mean PAPR for WGN, even when the spectrogram window overlap was as high as 75%. Further investigations, both theoretical and experimental, on the impacts of data correlations on the PAPR distribution for sampled I/Q data is an interesting topic for future work.

Most of the prior literature on PAPR has focused on using it as a figure of merit for the analog RF transmitter chain in a communication system. Specifically, there has been a lot of work on PAPR reduction schemes for OFDM communication signals [1] and power amplifiers to accommodate high PAPR [2]. On the other hand, there are few references that discuss PAPR in the context of digital signal analyzer measurements, with the exception of [3] and [22]. The present work helps to fill this gap by providing a simple, exact formula for the mean PAPR of WGN in I/Q data and showing how it can be applied to identify non-ideal I/Q acquisitions and to RF channel occupancy assessments.

ACKNOWLEDGMENTS

REFERENCES

[1] Y. Rahmatallah and S. Mohan, "Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1567–1592, 2013.

[2] G. Nikandish, R. B. Staszewski, and A. Zhu, "Breaking the bandwidth limit: A review of broadband Doherty power amplifier design for 5G," *IEEE Microwave Magazine*, vol. 21, no. 4, pp. 57–75, 2020.

[3] J. P. Dunsmore, *Handbook of Microwave Component Measurements*, 2nd ed. Hoboken, NJ: Wiley, 2020.

[4] M. Cotton and R. Dalke, "Spectrum Occupancy Measurements of the 3550–3650 Megahertz Maritime Radar Band near San Diego, California," National Telecommunications and Information Administration, Tech. Rep. TR-14-500, Jan 2014.

[5] D. Kuester *et al.*, "Radio Spectrum Occupancy Measurements Amid COVID-19 Telework and Telehealth," National Institute of Standards and Technology, Tech. Rep. TN-2240, Oct 2022.

[6] O. Stienne, V. Deniau, and E. P. Simon, "Assessment of transient EMI impact on LTE communications using EVM & PAPR," *IEEE Access*, vol. 8, pp. 227 304–227 312, 2020.

[7] J. G. Proakis and M. Salehi, *Communication Systems Engineering*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[8] G. Vasilescu, *Electronic Noise and Interfering Signals: Principles and Applications*. Springer, 2006.

[9] E. Milotti, *The Physics of Noise*. Morgan & Claypool Publishers, 2019.

[10] D. Wulich, "Comments on the peak factor of sampled and continuous signals," *IEEE Communications Letters*, vol. 4, no. 7, pp. 213–214, 2000.

[11] S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach*, 3rd ed. New York: McGraw-Hill, 2006.

[12] G. Wunder and H. Boche, "Peak value estimation of bandlimited signals from their samples, noise enhancement, and a local characterization in the neighborhood of an extremum," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 771–780, 2003.

[13] H. Boche and U. J. Mönich, "On the solvability of the peak value problem for bandlimited signals with applications," *IEEE Transactions on Signal Processing*, vol. 69, pp. 103–118, 2020.

[14] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer, 1983.

[15] P. H. Wirsching, T. L. Paez, and K. Ortiz, *Random Vibrations: Theory and Practice*. New York: John Wiley & Sons, 1995.

[16] S. O. Rice, "Mathematical analysis of random noise," *The Bell System Technical Journal*, vol. 23, no. 3, pp. 282–332, 1944.

[17] S. O. Rice, "Mathematical analysis of random noise," *The Bell System Technical Journal*, vol. 24, no. 1, pp. 46–156, 1945.

[18] D. E. Cartwright and M. S. Longuet-Higgins, "The statistical distribution of the maxima of a random function," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 237, no. 1209, pp. 212–232, 1956.

[19] H. Ochiai and H. Imai, "On the distribution of the peak-to-average power ratio in OFDM signals," *IEEE Transactions on Communications*, vol. 49, no. 2, pp. 282–289, Feb 2001.

[20] E. Castillo, *Extreme Value Theory in Engineering*. San Diego, CA: Elsevier, 1988.

[21] H. A. David, *Order Statistics*, 2nd ed. New York: Wiley, 1981.

[22] Keysight Technologies, Inc., "Spectrum and Signal Analyzer Measurements and Noise," Application Note, Aug 2021. [Online]. Available: https://www.keysight.com/us/en/assets/7018-06765/application-notes/5966-4008.pdf

[23] M. M. Siddiqui, "Some problems connected with Rayleigh distributions," *Journal of Research of the National Bureau of Standards*, vol. 66D, no. 2, pp. 167–174, March-April 1962.

[24] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 1*, 2nd ed. New York: Wiley, 1994.

[25] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Pacific Grove, CA: Duxbury, 2002.

[26] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, 2nd ed. Reading, MA: Addison-Wesley Publishing, 1994.

[27] M. S. Longuet-Higgins, "On the statistical distribution of the heights of sea waves," *Journal of Marine Research*, vol. 11, no. 3, pp. 245–266, 1952.

[28] P. Virtanen *et al.*, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[29] National Instruments, "PXIe-5646 Specifications," retrieved: 2024-06-20. [Online]. Available: https://www.ni.com/docs/en-US/bundle/pxie-5646-specs/page/specs.html

[30] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York: John Wiley & Sons, 1992.

[31] L. Smaini, *RF Analog Impairments Modeling for Communication Systems Simulation: Application to OFDM-based Transceivers*. Hoboken, NJ: John Wiley & Sons, 2012.

[32] M. A. Richards, "The discrete-time Fourier transform and discrete Fourier transform of windowed stationary white noise," Georgia Institute of Technology, Tech. Rep., 2013. [Online]. Available: https://sites.gatech.edu/wp-content/uploads/sites/462/2016/08/DFT-of-Noise.pdf

[33] Deepwave Digital, Inc., "AIR-T Technical Specifications," [Accessed: Sept. 6, 2024]. [Online]. Available: https://docs.deepwavedigital.com/AIR-T/Embedded_Series/#sdr-technical-specifications

[34] Federal Communications Commision, 2023, accessed: Sept. 6, 2024. [Online]. Available: https://www.fcc.gov/wireless/bureau-divisions/mobility-division/35-ghz-band/35-ghz-band-overview

[35] Federal Communications Commision, 2021, accessed: Sept. 6, 2024. [Online]. Available: https://www.fcc.gov/auction/107